

## MEMORY DEVICE AND METHOD HAVING DATA PATH WITH MULTIPLE PREFETCH I/O CONFIGURATIONS

### TECHNICAL FIELD

This invention relates to memory devices, and more particularly to a  
5 memory device data path and method that can operate in either a high-speed, narrow  
data bus mode or a low-speed, wide data bus mode.

### BACKGROUND OF THE INVENTION

Memory devices, such as dynamic random access memories  
("DRAMs"), have a variety of performance parameters. One of the most important of  
10 these performance parameters is the speed at which memory devices are able to read  
and write data. Generally, memory devices capable of reading and writing data at a  
higher speed, known as high performance memory devices, are more expensive.  
Conversely, memory devices that are only capable of accessing data at a slower rate,  
known as low performance memory devices, must be sold at a cheaper price. In an  
15 attempt to increase the operating speed of memory devices, double data ("DDR") rate  
DRAMs have been developed. DDR DRAMs are synchronous DRAMs that perform  
two memory operations each clock cycle—one on each transition of each clock pulse.  
In a typical DDR DRAM, the memory cells in two adjacent columns having the same  
column address are read each clock cycle.

20 Another performance parameter applicable to memory devices is the  
width of the memory device's data bus. Wider data buses operating at a given speed  
have a higher bandwidth, *i.e.*, a greater number of bits/second can be accessed. The  
data bus of most memory devices, such as DRAMs, generally have a width of various  
powers of 2, *i.e.*, 4, 8, 16, etc. bits.

25 The need to provide memory devices having different performance  
parameters generally requires memory device manufacturers to design and manufacture  
a wide variety of memory devices. For example, memory device manufacturers must  
design and fabricate relatively expensive memory devices that are capable of operating

at a high-speed and different, relatively inexpensive memory devices that are only capable of operating at a relatively low-speed. Unfortunately, it is expensive to design each memory device and the processing needed to fabricate the memory device. The expense of designing and fabricating a variety of different memory devices having  
5 different performance parameters is exacerbated by the rapid obsolescence of memory devices as newer devices are introduced at an ever faster rate.

There is therefore a need for memory devices, such as DRAMs, that are capable of operating as either high-speed, narrow data bus memory devices or a low-speed, wide data bus memory devices.

## 10 SUMMARY OF THE INVENTION

Data are coupled from a memory array to data bus terminals by transferring  $2N$  bits of parallel data from the array in a first mode and  $N$  bits of parallel data in a second mode. The parallel data are transferred from the array to parallel-to-serial converters using a bus having a width of  $N$  bits. The parallel-  
15 to-serial converters convert the parallel data bits to respective bursts of serial data containing  $N/M$  bits and apply the bursts to  $2M$  data bus terminals in the first mode and  $M$  data bus terminals in the second mode. The data may be transferred from the memory array in the first operating mode by transferring first and second sets of  $N$  data bits from the array in respective first and second read  
20 operations. Alternatively,  $2N$  data bits may be transferred from the memory array in a single read operation. As a result, data may be transferred to  $M$  data bus terminals at a relatively high-speed in a high performance mode, or to  $2M$  data bus terminals at a relatively low-speed in a low performance mode.

## BRIEF DESCRIPTION OF THE DRAWINGS

25 Figure 1 is a block diagram of a memory device in accordance with one embodiment of the invention.

Figure 2 is a block diagram of a memory array used in the memory device of Figure 1.

Figure 3 is a block diagram of one of several memory array mats used in the memory array of Figure 2.

Figure 4 is a block diagram of one of several memory sub-arrays used in the memory array mat of Figure 3.

5           Figure 5 is a block diagram of a portion of a data path used in the memory device of Figure 1.

Figure 6 is a logic and block diagram of one of several parallel-to-serial converters used in the portion of a data path shown in Figure 5.

10           Figure 7 is a block diagram of a computer system using the memory device of Figure 1.

#### DETAILED DESCRIPTION OF THE INVENTION

A memory device in accordance with one embodiment of the invention is illustrated in Figure 1. The memory device illustrated therein is a synchronous dynamic random access memory ("SDRAM") 10, although the invention can be embodied in  
15 other types of DRAMs, such as packetized DRAMs and RAMBUS DRAMs (RDRAMs"), as well as other types of memory devices, such as static random access memories ("SRAMs"). The SDRAM 10 includes an address register 12 that receives either a row address or a column address on an address bus 14. The address bus 14 is generally coupled to a memory controller (not shown in Figure  
20 1). Typically, a row address is initially received by the address register 12 and applied to a row address multiplexer 18. The row address multiplexer 18 couples the row address to a number of components associated with either of two memory banks 20, 22 depending upon the state of a bank address bit forming part of the row address. Associated with each of the memory banks 20, 22 is a  
25 respective row address latch 26 which stores the row address, and a row decoder 28 which applies various signals to its respective array 20 or 22 as a function of the stored row address. The row address multiplexer 18 also couples row addresses to the row address latches 26 for the purpose of refreshing the memory

cells in the arrays 20, 22. The row addresses are generated for refresh purposes by a refresh counter 30, which is controlled by a refresh controller 32.

After the row address has been applied to the address register 12 and stored in one of the row address latches 26, a column address is applied to the address register 12. The address register 12 couples the column address to a column address latch 40. Depending on the operating mode of the SDRAM 10, the column address is either coupled through a burst counter 42 to a column address buffer 44, or to the burst counter 42 which applies a sequence of column addresses to the column address buffer 44 starting at the column address output by the address register 12. In either case, the column address buffer 44 applies a column address to a column decoder 48 which applies various signals to respective sense amplifiers and associated column circuitry 50, 52 for the respective arrays 20, 22.

Data to be read from one of the arrays 20, 22 is coupled to the column circuitry 50, 52 for one of the arrays 20, 22, respectively. The data is then coupled through a read data path 54 to a data output register 56, which applies the data to a data bus 58. Data to be written to one of the arrays 20, 22 is coupled from the data bus 58 through a data input register 60 and a write data path 62 to the column circuitry 50, 52 where it is transferred to one of the arrays 20, 22, respectively. A mask register 64 may be used to selectively alter the flow of data into and out of the column circuitry 50, 52, such as by selectively masking data to be read from the arrays 20, 22.

The above-described operation of the SDRAM 10 is controlled by a command decoder 68 responsive to command signals received on a control bus 70. These high level command signals, which are typically generated by a memory controller (not shown in Figure 1), are a clock enable signal CKE\*, a clock signal CLK, a chip select signal CS\*, a write enable signal WE\*, a row address strobe signal RAS\*, and a column address strobe signal CAS\*, which the "\*" designating the signal as active low. Various combinations of these signals are registered as respective commands, such as a read command or a write

command. The command decoder 68 generates a sequence of control signals responsive to the command signals to carry out the function (e.g., a read or a write) designated by each of the command signals. These command signals, and the manner in which they accomplish their respective functions, are conventional.

- 5 Therefore, in the interest of brevity, a further explanation of these control signals will be omitted.

The read data path 54 from the column circuitry 50, 52 to the data output register 56 includes one or more pairs of complimentary input/output (“I/O”) lines (not shown in Figure 1) that couple data from a sense amplifier (not shown) for each column in each array 20, 22, respectively. The sense amplifier in the column circuitry 50, 52 for an addressed column receives complimentary signals from a pair of complimentary digit lines. The digit lines are, in turn, coupled to a pair of the complimentary I/O lines by column addressing circuitry. Each pair of I/O lines is selectively coupled by a pair of complimentary data lines to the complimentary inputs of a DC sense amplifier (not shown) included in the read data path 54. The DC sense amplifier, in turn, outputs data to the data output register 56, which is coupled to output or “DQ” terminals of the memory device 10. As explained in detail below, the SDRAM 10 according to one embodiment of the invention includes 16 DQ terminals, 8 of which are used in the high-speed mode and 16 of which are used in the low speed mode. Each of the DQ terminals coupled serial data to or from the DRAM 10 in bursts of 8 bits.

One of the memory arrays 20 is illustrated in Figure 2. The memory array 20 includes 8 memory array “mats” 100a-h divided into 4 banks, which are labeled in Figure 2 at B0-B3. However, it will be understood that the memory array mats 100a-h may be arranged in a greater or lesser number of banks, and the memory array 20 may contain a greater or lesser number of memory array mats 100. The read data path 54 (Figure 1) includes a first I/O bus 104 having 32 pairs of complimentary I/O lines coupled to the memory array mats 100a,b,e,f, and a second I/O bus 106 having 32 pairs of complimentary I/O lines coupled to the memory array mats 100c,d,g,h.

One of the memory array mats 100 used in the memory array 20 according to one embodiment of the invention is illustrated in Figure 3. The mat 100 includes 256 sub-arrays 110 arranged in 16 columns and 16 rows. Each of the memory mats 100 includes 16 column lines 114, each of which, when activated, selects the  
 5 corresponding column. The memory mat 100 also includes a large number of row lines (not shown), which, when activated, selects a respective row in the sub-arrays 110. A set of 4 flip-flops 120 is positioned beneath each column of the memory mat 100. When a row line is activated, 4 bits of data are coupled from the memory mat 100 from each column selected by activating a respective column line 114. The 4 bits of data for  
 10 each column are coupled from the memory mat 100 to a respective set of flip flops 120 through a respective digit line bus 122 that includes 4 complimentary digit lines. Thus, when 8 column lines 114 are activated, 32 bits of data are stored in 8 sets of flip-flops corresponding to the respective activated column lines 114.

As shown in Figure 4, each of the sub-arrays 110 includes 256k memory  
 15 cells (not shown) arranged in rows and columns. When a row of the memory mat 100 is activated and a column line 114 is selected, 4 complimentary digit lines 130 in 4 respective columns of the sub-array 110 are coupled to 4 respective flip-flops 120. The flip-flops 120, in turn, drive respective complimentary pairs of I/O lines 140. In operation, 8 columns of each memory mat 100 are activated at a time, so that the 8 sub-  
 20 arrays 110 in 8 respective active columns each output 4 bits of data. Each memory array mat 100 thus provides 32 bits of data, which are temporarily stored in the flip-flops 120. Since two memory array mats 100 are used for each bank, each bank B0-B3 outputs 64 bits of data. In operation, the 4 data bits coupled from each sub-array 110 are prefetched and stored in the flip-flops 120 for subsequent coupling to the DQ  
 25 terminals (Figure 1), as explained in greater detail below.

The data bits are transferred from the flip-flops 120 in either of two modes, depending on whether the memory device 10 is operating in either the high-speed mode or the low-speed mode. In the high-speed mode, 8 bits of data stored in respective flip-flops 120 are transferred serially to a respective data bus (DQ) terminal.  
 30 The manner in which the parallel data stored in the flip-flops 120 are converted to serial

data will be explained with reference to Figures 5 and 6. Since there are 32 bits stored in respective flip-flops 120 for each memory array mat 100, the 32 bits are coupled in serial bursts of 8 bits to each of 4 data bus terminals in the high-speed mode. The 32 bits stored in the flip-flops 120 for the other memory array mat 100 are also coupled in serial bursts of 8 bits to each of 4 data bus terminals. As a result, in the high-speed mode, 64 bits are coupled in serial bursts of 8 bits to each of 8 data bus terminals.

In the low-speed mode, the data bits stored in the flip-flops 120 are also transferred serially to a respective data bus (DQ) terminal. However, in the low-speed mode, the data bits are transferred to 16 data bus terminals. Yet circuitry (not shown) interfacing with the memory device 10 is adapted to receive data in bursts of 8 bits from each of the 16 data bus terminals. Thus, in the low-speed mode, 128 bits are required to couple bursts of 8 bits to each of 16 data bus terminals. Since there are 32 bits stored in each set of respective flip-flops 120 for each memory array mat 100, the 64 bits stored in the flip-flops 120 for both memory array mats 100 can supply only half of the required number of data bits. As a result, in the low-speed mode, two sets of 64 bits must be prefetched and stored in the flip-flops 120 before the prefetched data bits can be coupled to the data bus terminals. The reason this operating mode is considered a low-speed mode is because of the extra time needed to prefetch and/or coupled twice as many data bits from each memory array mat 100 in the low-speed mode compared to the high-speed mode. Therefore, to prefetch 64 bits from each memory array mat 100, the memory device must perform 2 read operations with each data transfer, which requires substantially more time than a single read operation. However, the bandwidth of the memory device 10 is somewhat the same in both modes. In the high-speed mode, twice as many memory devices 10 are needed to provide data to the 16 data bus terminals compared to the low speed mode, but the data is provided twice as fast.

To help maintain the operating speed of the memory device in the low-speed mode, the memory device 10 may, instead of performing two read operations to provide 128 bits of data, simultaneously activate all 16 columns in each memory array mat 110. Thus, each memory array mat 110 prefetches 64 bits (4 bits from each

column) during each read operation. As a result, all 64 of the flip-flops 120 shown in Figure 3 are needed for each memory array mat 110 to store the 64 prefetched bits.

In operation, 4 data bits prefetched from each column and stored in respective flip-flops 120 are coupled through a respective pair of I/O lines 140 during the first part of each read cycle, and 4 data bits prefetched from another column and stored in respective flip-flops 120 are coupled through the same pair of I/O lines 140 during the second part of each read cycle. Thus, in this alternative low-speed mode, 128 bits of data stored in respective flip-flops 120 for both memory array mats 110 are coupled through 64 pairs of complimentary I/O lines for each read operation. In contrast, in the high-speed mode described above, 64 bits of data stored in respective flip-flops 120 for both memory array mats 110 are coupled through 64 pairs of complimentary I/O lines. As a result, in the low-speed mode, twice as many data bits must be coupled through the data lines during the same period of time. It is for this reason, that this operating mode is considered a low-speed mode even though it does not require 2 read operations for every read cycle.

The manner in which prefetched data bits are coupled between the flip-flops 120 and data bus terminals for one of the memory mats 100 is shown in Figure 5. The circuitry shown in Figure 5 is adapted to be used with the first embodiment of the low-speed operating mode in which two reads are performed for every read operation. However, it will be understood that the circuitry can be easily modified for the alternative embodiment in which every column of each memory array mat 100 is read and twice as many flip-flops 120 are provided.

With reference to Figure 5, the prefetched 32 data bits stored in the respective flip-flops 120 are coupled through 32 respective complimentary pairs of I/O lines 140. Eight groups of 4 I/O line pairs 140 are coupled to 8 respective parallel-to-serial converters 150 so that 4 pairs of I/O lines 140 are coupled to each converter 150. However, 4 of the converters 150a include only 4 pairs of input lines, which are coupled to 4 pairs of I/O lines 140 of a respective group. The remaining 4 converters 150b include 8 input lines, which are coupled to 4 pairs of I/O lines 140 of a respective group and 4 pairs of I/O lines 140 that are coupled to one of the 4-input converters 150a.



In the low-speed mode, 4 bits of parallel data are coupled to each of the 16 converters 150a,b for each read operation, so that, after two read operations have been performed, 8 bits have been coupled to each of the 16 parallel-to-serial converters 150. The converters 150 then each output an 8-bit burst through respective I/O paths 134 to 16 respective data bus terminals 160. In the high-speed mode, 8 bits of parallel data are coupled to each of the four 8-input converters 150b, and the converters 150b then each output an 8-bit burst through I/O paths 134 to 8 respective data bus terminals 160. Thus, in the high-speed mode the four 4-input converters 150a and the data bus terminals 160 to which they are coupled are not used.

For a write operation, burst of 8 bits are applied to each of either 8 or 16 data bus terminals, depending upon whether the SDRAM 10 is operating in either the high-speed mode or the low-speed mode, respectively. Respective serial-to-parallel converters 168 then convert the 8-bit burst to either an 8 bits of parallel data (in the high-speed mode) or two sets of 4 bits of parallel data (in the low-speed mode). The 4 data bits applied to each column of the memory mat are then coupled to respective columns of each sub-array 110 in a write operation.

One embodiment of the 8-bit parallel-to-serial converters 150a is shown in Figure 6. As previously explained, the parallel-to-serial converter 150a is adapted to receive 8 bits of parallel data and output a burst of 8 serial bits. However, the 4-bit parallel-to-serial converters 150b are substantially identical, as explained further below. When parallel data are to be transferred from the flip-flops 120 to the converter 150a RinPar signal transitions high, thereby triggering a load logic circuit 162. The load logic circuit 162 then outputs a high Data Load 0 ("DatLoad0") output, which is applied to a 4 input latch 164. The latch 164 has a 4-bit parallel data input that is selectively coupled to 8 of the flip-flops 120. Thus, each data input of the parallel-to-serial converter 150a is coupled to the outputs of two flip-flops 120. The outputs of 4 of the flip-flops 120 are coupled to respective data input terminals on the low-to-high transition of the RinPar signal. The 4 bits of parallel data are then stored in the latch 164.

When the 4 bits of data stored in the latch 164 are to be shifted out of the latch, an Rin signal transitions high, thereby causing an inverter 168 to output a low to a NAND gate 170, which, with NAND gate 174, forms a set-reset flip-flop 176. The flip-flop 176 is then set, thereby causing the NAND gate 170 to output an active high Serial Unload ("SerUld") signal to the latch 164. The high SerUld signal causes the latch 164 to output an active low Busy signal, and, in response to a Serial Clock ("SerClk") signal, shift the 4 bits of stored data out of the latch one bit at a time on each SerClk transition.

The serial data at the output of the latch 164 is applied to a multiplexer 180. As explained further below, the output of an inverter 182 is initially low and the other input to the multiplexer 180 is initially high. As a result, the multiplexer 180 couples a 4-bit burst of serial data from the latch 164 to a double-edge triggered flip-flop 184 that is clocked by the SerClk signal and its complement. Thus, on a transition of the SerClk signal having one polarity, each bit of data is shifted into the double-edge triggered flip-flop 184, and that bit of data is then shifted out of the flip-flop 184 on the next transition of the SerClk signal having the opposite polarity.

The high Busy signal at the output of the latch 164 causes a NAND gate 190 to output a high to a clocked driver 194 that is identical to the double-edge triggered flip-flop 184. Thus, on each transition of the SerClk signal, the driver 194 outputs a high Tri-State ("TS") signal. The TS signal is used to switch circuitry (not shown) receiving the serial data from the flip-flop 184 downstream in the read data path 54 (Figure 1) from a tri-state (high impedance) to an active, low impedance state.

Prior to the end of 4 cycles of the SerClk, a second set of 4 flip-flops 120 are coupled to respective Data<0:3> terminals and the Rin signal transitions low. The high-to-low transition of the Rin signal causes the load logic circuit 162 to output a high Data Load 1 ("DatLoad1") output, thereby causing a second 4-input latch 200 to store the 4 bits of parallel data from the flip-flops 120.

When the 4 bits of data stored in the latch 164 have been shifted out responsive to 2 periods of the SClk signal, the latch 164 outputs a low DoneSync signal. The low DoneSync signal is applied to the NAND gate 174 to reset the flip-flop 176,

thereby causing the NAND gate 170 to disable the latch 164 from further outputting further serial data. The low DoneSync signal is also applied to a Done0 input of the load logic circuit 162 to subsequently allow the counter to be reset, as explained further below. Finally, the Done0 signal is applied to a NAND gate 204 that, with a NAND  
 5 gate 206, forms a flip-flop 208 that is set by the Done0 signal. When the flip-flop 208 is set, it causes the NAND gate 206 to output a high, which causes an inverter 210 to apply an active high signal to a Serial Unload ("SerUld") terminal of the latch 200. The latch 200 then applies the 4 stored bits to the multiplexer 180 responsive to the SerClk signal, as explained above for the latch 164. The active high signal applied to the  
 10 SerUld terminal of the latch 200 also causes the latch 164 to apply an active low Busy signal to the NAND gate 190. The NAND gate 190 then applies a high to the driver 194 to cause the NAND gate 190 to output a high TS signal, as previously explained.

Setting the flip-flop 208 also causes the NAND gate 204 to output a low, which causes an inverter 214 to apply a high to the inverter 182 and the multiplexer  
 15 180. The multiplexer 180 then couples the output of the latch 200 to the double-edge triggered flip-flop 184.

When the 4 bits of data stored in the latch 200 have been shifted out of the latch 200, the latch 200 outputs a low DoneSync signal. The low DoneSync signal is applied to the NAND gate 200 to reset the flip-flop 208, thereby disabling the latch  
 20 200 from further outputting serial data. The low DoneSync signal is also applied to a Done1 input of the load logic circuit 162 to reset the load logic circuit 162 in preparation for a subsequent transition of the RinPar signal.

The flip-flops 176, 208 and the load logic circuit 162 can also be reset by an active low Reset signal, which is typically generated upon power-up.

25 As previously mentioned, the parallel-to-serial converter 150a converts 2 loads of 4 parallel bits of data to a burst of 8 serial bits of data. The parallel-to-serial converter 150a can easily be modified to implement the converter 150b that converts 4 either 4 or 8 parallel bits of data to a burst of 8 serial bits of data. For example, the converter 150b can be implemented by simply adding a set of 4 latches (not shown) to  
 30 latch the parallel load of 8 bits. These latches are disabled when in the 4-bit mode and

the converter 150b operates in the same manner as the converter 150a as described above.

Figure 7 shows a computer system 300 containing the SDRAM 10 of Figure 1. The computer system 300 includes a processor 302 for performing various computing functions, such as executing specific software to perform specific calculations or tasks. The processor 302 includes a processor bus 304 that normally includes an address bus, a control bus, and a data bus. In addition, the computer system 300 includes one or more input devices 314, such as a keyboard or a mouse, coupled to the processor 302 to allow an operator to interface with the computer system 300. Typically, the computer system 300 also includes one or more output devices 316 coupled to the processor 302, such output devices typically being a printer or a video terminal. One or more data storage devices 318 are also typically coupled to the processor 302 to allow the processor 302 to store data in or retrieve data from internal or external storage media (not shown). Examples of typical storage devices 318 include hard and floppy disks, tape cassettes, and compact disk read-only memories (CD-ROMs). The processor 302 is also typically coupled to cache memory 326, which is usually static random access memory ("SRAM"), and to the SDRAM 10 through a memory controller 330. The memory controller 330 normally includes a control bus 336 and an address bus 338 that are coupled to the SDRAM 10. A data bus 340 is coupled from the SDRAM 10 to the processor bus 304 either directly (as shown), through the memory controller 330, or by some other means.

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. For example, although the SDRAM 10 has 16 data bus terminals, 16 DQ terminals, 8 of which are used in the high-speed mode and 16 of which are used in the low speed mode, it will be understood that memory devices may have a lesser or greater number of DQ terminals. Also, each burst of data may contain a lesser or greater number of bits than the 8-bit bursts described herein, and the width of the I/O path coupling data between the memory array and the parallel-

to-serial converters may be wider or narrower than the I/O path described herein. Other variations will also be apparent to one skilled in the art. Accordingly, the invention is not limited except as by the appended claims.